Chapter 14

Comparative Cognitive Task Analysis

Susan S. Kirschenbaum
Naval Undersea Warfare Center Division Newport, Newport, RI

J. Gregory Trafton
Naval Research Laboratory, Washington, DC

Elizabeth Pratt
University of Connecticut

It is easy to force a weather forecaster to work out of context—simply move him or her to some new locale. It takes at least one full seasonal cycle for forecasters to reacquire expertise. Worse, move a forecaster from the Northern Hemisphere to the Southern Hemisphere. Major things change. Low-pressure systems spiral clockwise, not counterclockwise. Effects of ocean currents, seasonal variations, and effects of land masses change everything. Any knowledge of trends that the forecaster had relied on are now utterly useless.

In the studies we report in this chapter, we did something like this, but the switch involved making *us*, as cognitive systems engineers, work out of context. Work on forecaster reasoning with which we are familiar (e.g., Hoffman, 1991), including on our own research, has involved the study of forecasters in the U.S. Navy and U.S. National Weather Service. We think we have some ideas about how forecasters think (see chap. 15, this volume), but are we sure? How does what we *think* we know transfer to, say, forecasting in Australia? Or does it transfer? What if we were to advise building some new tool, only to learn that it does not help forecasters in regions other than the continental United States?

Weather forecasting is a complex process. The supporting information is multidimensional, distributed, and often uncertain. It includes both "raw" observations (e.g., current temperature, winds, pressure, clouds, precipitation, radar returns, satellite pictures, etc.) and analytic weather models that predict future weather conditions at various scales of space and time. The

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **2007**	2. REPORT TYPE	3. DATES COVERED **00-00-2007 to 00-00-2007**

4. TITLE AND SUBTITLE **Comparative Cognitive Task Analysis**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Naval Research Laboratory,Navy Center for Applied Research in Artificial Intelligence (NCARAI),Code 5510, 4555 Overlook Ave., SW,Washington,DC,20375**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR’S ACRONYM(S)
	11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**10**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

information that the weather forecaster uses is often downloaded from external Web sites. Local weather organizations use (or build) support tools for displaying downloaded data and images and for building and displaying their own forecasts.

To optimize these tools, consideration must be given to the user-tool-task triad that is central to the principles of human-centered computing (HCC) (Hoffman, Coffey, Ford, & Bradshaw, 2001; Hoffman, Hayes, Ford, & Hancock, 2002). These principles require the designer to build tool that facilitate the task that the user does and accommodate human perceptual, cognitive, and motor functioning. How does the designer incorporate the user-tool-task triad of HCC into this complex and specialized domain? How does the designer gain enough knowledge of the users' tasks and processes to provide useful assistance? And how does the designer disentangle the effects of task, training, teamwork arrangements, and basic human cognition from those of the design of the tools?

The traditional way human factors engineers approach this problem is to perform a task analyses to determine how people operate in a specific domain on a specific task. Cognitive Task Analysis (CTA) is a set of methods that takes into account the perception (i.e., vision), cognition (i.e., decision making), and motor actions (i.e., mouse movements) needed to accomplish a task. In this chapter, we build on CTA methods by suggesting that comparative cognitive task analysis (C2TA) can help solve the aforementioned problems. C2TA is based on replication studies conducted in different environments. Replication is a basic principle of the scientific method, but usually replication aims at duplicating the original conditions. Comparative studies are also a common scientific practice. Within the HCC literature, comparative studies usually employ a traditional experimental design to ask such questions as which device or design is faster (Haas, 1989), more efficient (Haas, 1989), and/or lowers workload (Kellogg & Mueller, 1993). However, CTA is often an exploratory research strategy that focuses on process rather than final performance (Sanderson & Fischer, 1994). C2TA draws on all these traditions, applying elements of replication and comparative methods to the exploratory process approach of CTA. Because it derives data from more than one environment, C2TA provides insight into interface design that single-site studies and individual CTA methods cannot.

There are many versions of task analysis ranging from time-and-motion study (Gilbreth & Gilbreth, 1917) to GOMS (goals, operators, methods, selection rules) analysis (Card, Moran, & Newell, 1983), to ecological interface design (EID) (Vicente & Rasmussen, 1992). Each is best suited to particular aspects of design problems. For example, GOMS analysis is a keystroke-level process for describing human–computer interactions (e.g., mouse and keyboard interactions). EID focuses on how the operator interacts with indicators of physical functioning such as in a power plant or manufacturing control room. CTA is especially useful in situations where the task is heavily dependent on human interpretation and integration of dynamic and highly uncertain data (Schraagen, Chipman, & Shalin, 2000).

Weather forecasters typically deal with large amounts of data over time and space (Hoffman, 1991). Additionally, the information they examine is uncertain on several dimensions (i.e., the predictive weather models that are run may be based on a small number of data points in some areas—like in the middle of the ocean—which necessitates interpolating from the current data, which may cause the final output to be more uncertain). The need for expertise at interpreting the weather prediction models, the dynamic nature of weather, and the uncertainty in the weather models makes weather forecasting an excellent candidate for CTA.

However, most of the data analyzed by CTA methods come from a single source (i.e., most CTA studies have been performed on a single system and/or a small group of people). Although the single approach is adequate in many situations, it may not be as generalizable as it could be. That is, any problems might be traced to the interaction between the person and the system. You may discover, for example, that a specific pointing device is not very effective on a particular system, but you do not know if that is a limitation of the pointing device or the way people (in general) think, or the way people in a particular organization think; you only know that the combination of people and pointing device on the task you are examining is not very effective. By examining different tools (i.e., different types of pointing devices on similar tasks), you can start to dissociate the effects of cognition and those of the tool.

For example, the pen, the typewriter, and the computer keyboard are all tools that can be used for writing a document. The writing process consists of planning, composing, editing, and production (writing/typing). The quantity and sequence of these processes is differentially supported by the three tools. The computer supports longer compositions, however, the writer plans longer before editing with a pen (Haas, 1989). This may be because editing with a pen includes crossing out, rewriting, cutting pages apart and taping them back together, arrows for inserts, and so on, and then repeating the production process (rewriting) on clean paper. Editing on a typewriter uses similar cross-out, cut, glue, and retype processes. With both of these tools, the rewrite (production) process is effortful. However, writers using a computer edit more as they write and new versions do not require redoing the physical production (Kellogg & Mueller, 1993).

The data for the two analyses reported here were collected during two studies in two different locations, a United States Navy (USN) Meteorology and Oceanography (METOC) center in California and a Royal Australian Navy (RAN) METOC facility. These studies employed the methods of cognitive field research and quasi-naturalistic observation in what Hoffman and Deffenbacher (1993) termed a "laboratory field" study. The studies were part of a project to provide improved tools for Navy weather forecasting. Only by understanding current practices and forecasting tools could improvements be suggested that would make the process more efficient while retaining accuracy levels. (Because accuracy data are regarded as sensitive, they are not reported here.) The two studies allowed us to map the information usage of decision makers to information visualization tools, and to compare the USN and RAN forecasters in order to distinguish between effects that are dictated by the tools and training of these specialists and those due to basic human cognition.

The intent of this chapter is to introduce a new approach to answering the questions in the previous paragraph. We need not report a full analysis of the data, but we do present sample results. We first briefly describe the data collection at the two sites. Then we review the results of the C2TA and show how suggestions for the design or redesign of tools flow from the C2TA results. More detailed results from both studies can be found in Kirschenbaum (2002) and Trafton et al. (2000)

THE TWO STUDIES

Study 1: U.S. Navy, 2000

Study 1 took place in San Diego, California, at a Naval meteorological and oceanographic facility. We set up a simulated METOC center with computer access to the tools that the forecasters typically use. Most of the weather information came from meteorological Web sites including military, nonmilitary government, and university sites.

Three pairs of participants consisting of a forecaster and a technician took part in the study. Each pair developed a forecast and prepared a forecast briefing for a (pretend) air strike to take place 12 hours in the future on Whidbey Island, Washington. All actions were videotaped and the participants were requested to "talk aloud" so as to produce a verbal protocol.

Study 2: Royal Australian Navy, 2001

The second study was a naturalistic observation of RAN forecasters working at a Weather and Oceanography Centre at an airbase in eastern Australia. Like their USN counterparts, they were forecasting for 12-, 24-, and 72-hour intervals for air operations. They prepared forecasts and forecast briefings, and used computer-based tools. As with the USN forecasters, most of the forecasting information came from meteorological Web sites. Also as in our study of the USN forecasters, they were videotaped and instructed to "talk aloud" to produce verbal protocols.

By retaining the task (forecasting) and moving to another group of practitioners with different tools (workstations, software), training, and teamwork practices, we might disentangle the effects due to human cognition, versus those due to the organizations, versus those due to the tools used—thereby permitting inferences about how to better support the common forecasting tasks for both groups.

RESULTS

The data were analyzed at two levels of detail. The first is a high-level description of the stages of weather forecasting. The second is a detailed description of the information-processing procedures used during each stage.

Information Use

Comparative CTA can tell two kinds of stories. *Similarities* in classes of information usage that are independent of the tools, training, and teamwork patterns imply basic processes of human cognition. In contrast, we can impute *differences* in information usage patterns as being due to the impact of differences in tools, training, and teamwork. To find either, we must code the verbal protocols to capture the way the forecasters use information. To analyze these data we selected usage encodings that capture what the forecaster did with the information. In other reports, we have examined the format of the information (text, graph, animation, etc.) or the form of the information (qualitative or quantitative) (Kirschenbaum, 2002; Trafton et al., 2000).

The major encoding categories for cognitive activities that we used are described in Table 14.1. Note that, in terms of expertise required and cognitive effort, there is a clear ordering from simplest to most demanding: Record <Extract < Compare < Derive.

The transcripts were encoded using the Table 14.1 categories, and the results for the USN and RAN forecasters were compared. Overall, the results indicated a strong similarity between USN and RAN information usage—the basic processes are the same. There were no methods that were used by one group but not by the other. However, the order, tools used, and relative frequency with which these methods were used did show significant differences in some areas. These areas are indications that the tools differentially support the tasks. They are of interest for C2AT and for the information they provide about opportunities to improve the toolset.

TABLE 14.1
Categories of Cognitive Work Used in the Analysis

Action	*Definition*	*Example*
Extract	To read information from any visible source. This occurs when a forecaster examines a visualization and extracts some sort of local or global features that are explicitly represented in the visualization.	"Looks like PVA over the area."
Compare	To use two or more sources and comparing them on any data.	"Radar shows precipitation, but I can't really see anything on the satellite picture."
Derive	To combine information that is available in the visualizations with the forecaster's own knowledge, so as to make inferences and come to a conclusion that differ from what is in the visible source.	"I think that's probably a little fast due to the fact that I don't think the models taking into account the topography of the area."
Record	To write down or copy information for reporting to users. It need not be the final form.	"This is a good picture right here, I'll take this. . . . Just crop this picture a little bit."

Note. The examples come from USN transcripts.

Figure 14.1 indicates differences in the details of how USN and RAN forecasters accomplish their task, using the resources at hand and within their own specific environments (weather, training, and manning). We concentrate on differences during the central tasks of developing and verifying the forecast. (There are no differences in the relative frequency of *record* actions even though specific tools and the pattern of tasks did differ.)

Two observations stand out. The RAN forecasters appear to spend the same proportion of time in *extracting, comparing,* and *deriving* information whereas the USN forecasters spend significantly more of their time extracting information, $\chi^2(3) = 31.31$, $p < .001$. In contrast, RAN forecasters spend virtually as much time *comparing* as *extracting* data. Thus, compared to the USN forecasters, the RAN forecasters spent a significantly larger proportion of their time engaged in comparing information, $\chi^2(1) = 7.28$, $p < .01$.

C2TA reveals the differences between the two groups. However, the analyst must find the reasons for these differences. Candidate causes include task, tool, and training differences. In this case, the goal is the same, predicting weather for naval aviation operations in the 12+-hour time frame. Though training differs between the groups, tool differences appear to be the more likely cause. For example, the RAN forecasters have better support for comparisons because they either use adjacent monitors or adjacent windows on the same monitor. Thus, they can see a satellite or radar pic-

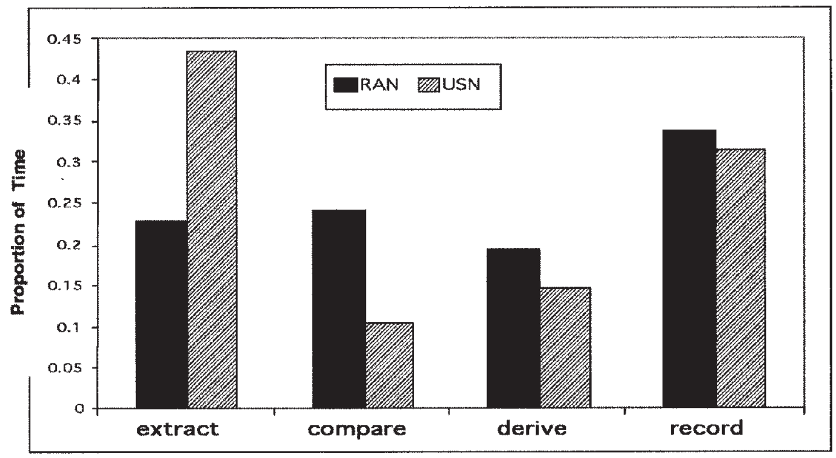

FIG. 14.1. Proportion of time spent performing cognitive operations.

ture simultaneously or can examine the outputs of two or more computer forecasting models side by side on the same monitor, as shown in Figure 14.2. In contrast, the USN forecasters must extract information from one data source, and then either compare it to information shown at some other workstation, or store it in memory or on paper, and then make comparisons from memory. With the RAN dual view (either on the same or adjacent monitors) the forecaster can make direct *comparisons.* The process of extraction is an integral part of the process while the storage burden is greatly reduced. In Figure 14.2, the forecaster is comparing the outputs of two computer models displayed side by side on the same monitor. Other comparisons observed were comparisons of predictions for the same computer model across time and comparisons of the computer model prediction for current time and current observations (e.g., a satellite image on an adjacent monitor).

Sequences

Another observation from Figure 14.1 is that both groups spend a considerable portion of their time recording information for use in their forecasts. Further insight into this process can be achieved by examining the sequence of processes. Table 14.2 shows the probability of going from one process to another for a USN and a RAN forecaster. For example, given that the RAN forecaster is currently extracting data, the probability of his next action being *comparing, deriving,* or *recording* are $p = .11$, .44, and .44, respectively. This transition table emphasizes the importance of the two poles, *extract* and *record.* These are the most common transition points for both the

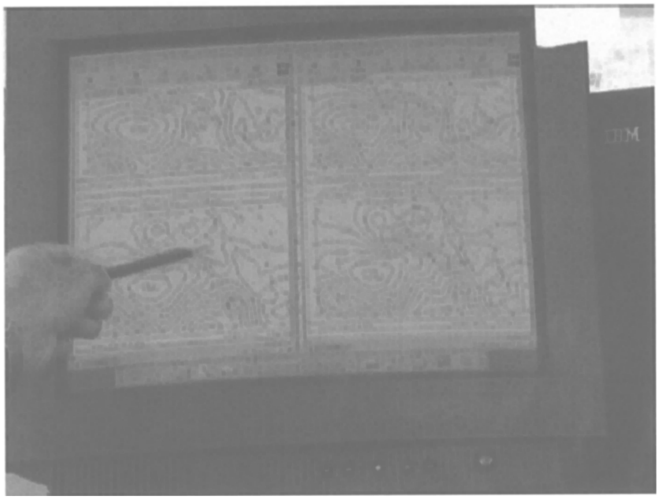

FIG. 14.2. Forecaster comparing the outputs of two computer models, using screen sectoring.

TABLE 14.2
Representative Transition Probabilities

From \ To	*Extract*	*Compare*	*Derive*	*Record*
RAN				
Extract	.00	.11	.44	.44
Compare	.40	.00	.20	.40
Derive	.31	.08	.00	.62
Record	.48	.24	.29	.00
USN				
Extract	.08	.25	.42	.25
Compare	.33	.00	.33	.33
Derive	.50	.13	.00	.38
Record	.60	.00	.20	.20

USN and RAN forecaster. Of the three-node transitions, the most common cycles for both were either

extract ™ *record* ™ *extract*

or

record ™ *extract* ™ *record.*

For RAN,

extract ™ *derive* ™ *record*

was also common. Transitions between *compare* and *derive* were noticeably fewer than those involving the poles.

As with the frequency data, sequence data provide insight into how tools do (or do not) support the cognitive tasks that make up weather forecasting. Design implications from the sequence data suggest the most effective places to automate. For example, as *extract* ™ *record* sequences are common, a semiautomated tool might allow the forecaster who is *extracting* information to *record* the selected data at the press of a button and without having to change screens. This would speed the recording process, eliminate accidental recording errors (typos, memory errors, etc.) and reduce the need to cycle between two tools.

IMPLICATIONS

C2TA is only one of the inputs to inform human-centered tool design. It is, however, necessary because without it, the tools would likely not meet the needs of the user, even if the designer were knowledgeable about the users' tasks. In contrast, with a "traditional" CTA, we could have observed the processes of *extraction, comparison, deriving,* and *recording* during the development of a weather forecast, and studied forecasters' cycle between developing their forecast (*extract, compare, derive*) and recording data. We would not have known whether these processes and cycles are common to other forecasting environments. Furthermore, we would not have learned the important role that the supporting tools play in the *comparison* process.

The cognitive systems engineer and the system developers and designers must work together to exploit these observations to guide the development of better tools. C2TA is only the first step but one that can inform and guide design toward making improvements where they are most needed. Our results suggest a need for better tools to further facilitate the *comparison* process, thus affirming an hypothesis about workstation design from traditional task analyses (Hoffman, 1991) that forecasters have to be able compare multiple data types at a glance. New display approaches and products are coming along to further support forecasting. For instance, it is now possible now to compare the outputs of the differing computer forecasting models. It is possible to superimpose computer model outputs over satellite pictures for current model comparisons.

These are just examples of the kinds of conclusions that can be derived from C2TA and have contributed to the momentum to develop better tools to help forecasters. With a single data set, the designer cannot know if the observed behavior is due to some demand characteristic of the tool set or to some facet of human cognition. With the addition of a second data set, the designer can separate the two and is thus free to develop better ways to support common cognitive processes with new tools.

ACKNOWLEDGMENTS

Susan Kirschenbaum's research was supported in part by the Office of Naval Research (ONR) as Project A424102 and by the Naval Undersea Warfare Center's (NUWC) In-House Laboratory Independent Research (ILIR) Program, as Project E102077. The ILIR program is funded by ONR. Greg Trafton's research was supported by Grants N00014-00-WX-20844 and N00014-00-WX-40002 from the Office of Naval Research.

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Navy.

REFERENCES

Card, S. K., Moran, T. P., & Newell, A. (1983). *The psychology of human–computer interaction.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Gilbreth, F. B., & Gilbreth, L. M. (1917). *Applied motion study.* New York: Macmillan.

Haas, C. (1989). Does the medium make a difference? Two studies of writing with pen and paper and with computers. *Human–Computer Interaction, 4,* 149–169.

Hoffman, R. R. (1991). Human factors psychology in the support of forecasting: The design of advanced meteorological workstations. *Weather and Forecasting, 6,* 98–110.

Hoffman, R. R., Coffey, J. W., Ford, K. M., & Bradshaw, J. (2001, September). Human-centered computing: Sounds nice, but what is it? Paper presented at the Human System Integration Symposium, Crystal City, VA.

Hoffman, R. R., & Deffenbacher, K. A. (1993). An analysis of the relations of basic and applied science. *Ecological Psychology, 5,* 315–352.

Hoffman, R. R., Hayes, P., Ford, K. M., & Hancock, P. A. (2002, May/June). The triples rule. *IEEE: Intelligent Systems,* pp. 62–65.

Kellogg, R. T., & Mueller, S. (1993). Performance amplification and process restructuring in computer-based writing. *International Journal of Man–Machine Studies, 39,* 33–49.

Kirschenbaum, S. S. (2002). Royal Australian Navy meteorology and oceanography operations research report (Tech. Rep. No. 11-346). Newport, RI: Naval Undersea Warfare Center Division.

Sanderson, P. M., & Fisher, C. (1994). Exploratory sequential data analysis: Foundations. *Human–Computer Interaction, 9,* 251–317.

Schraagen, J. M., Chipman, S. F., & Shalin, V. L. (Eds.). (2000). *Cognitive task analysis.* Mahwah, NJ: Lawrence Erlbaum Associates.

Trafton, J. G., Kirschenbaum, S. S., Tsui, T. L., Miyamoto, R. T., . Ballas, J. A., & Raymond, P. D. (2000). Turning pictures into numbers: Use of complex visualizations. *International Journal of Human–Computer Studies* [Special issue: *Empirical Evaluations of Information Visualization], 53,* pp. 827–850.

Vicente, K. J., & Rasmussen, J. (1992). Ecological interface design: Theoretical foundations. *IEEE Transactions on Systems, Man, and Cybernetics, 22,* 586–606.